

A New Day for Website Archiving 2.0

Jonathan Band

jband@policybandwidth.com

In October 2006, I wrote a memorandum explaining that two district court decisions, *Field v. Google*, 412 F. Supp. 2d 1106 (D. Nev. 2006), and *Parker v. Google*, 422 F. Supp. 2d 492 (E.D. Pa. 2006), provided several viable legal theories permitting libraries to engage in website archiving – in particular, fair use.¹ The memorandum concluded that these two district court opinions from different circuits provided a solid basis for arguing that copyright law permits libraries to archive websites.

Developments in the fair use jurisprudence since 2006 confirm this conclusion. A central issue in the fair use analysis is whether the use is “transformative.” *Campbell v. Acuff-Rose*, 510 U.S. 569, 579 (1994). In the website archiving context, the question is whether a library’s reproduction and subsequent display of entire websites without material alteration is “transformative.” The case law and legal opinions discussed below all indicate that library website archiving for the purpose of preservation and scholarship is transformative as that term is used by courts in the fair use context.

I. *Perfect 10 v. Amazon.com* (2007)

Perfect 10 published erotic photographs in a magazine and a website. It claimed that other websites copied and displayed its photographs without permission. In the course of its search engine operations, Google automatically scanned the photographs on the infringing websites, stored them in its search database, displayed low-resolution thumbnails of these infringing images in response to search queries, and provided links to the infringing sites.

Perfect 10 sued Google both for displaying thumbnail images of Perfect 10 photographs in response to search queries and for linking to sites where infringing images were displayed. Perfect 10 filed a motion for preliminary injunction, which the district court granted.² In its ruling on the display of the thumbnail images in the search results, the district court distinguished *Kelly v. Arriba Soft Corporation*, 336 F.3d 811 (9th Cir. 2003), where the Ninth Circuit held that a search engine’s storage and display of thumbnail images was a fair use. The district court found that Google’s use was “more commercial than” Arriba Soft’s because Google received a financial benefit from the display of the Perfect 10 thumbnails. Moreover, the district court found that Google’s display of the Perfect 10 thumbnails might harm an emerging market for the display of images of naked woman on mobile phones. Because of these factors, the district court concluded that Google was unlikely to prevail on its fair use defense.

The U.S. Court of Appeals for the Ninth Circuit reversed the district court’s rejection of Google’s fair use defense.³ The Ninth Circuit rejected the district court’s distinguishing of *Arriba Soft* on the basis of Google’s service being more commercial than Arriba Soft’s. Further, the Ninth Circuit found that there was no evidence that the Google thumbnails superseded the mobile-phone download market.

At the same time, the court held that Google’s use of the thumbnails was “highly

¹ The 2006 memorandum is attached in the Appendix.

² *Perfect 10 v. Google*, 416 F. Supp 2d 828 (D. Cal. 2006), *aff’d in part, rev’d in part*, 508 F.3d 1146 (9th Cir. 2007).

³ *Perfect 10 v. Amazon.com*, 508 F.3d 1146 (9th Cir. 2007). The case name is styled *Perfect 10 v. Amazon.com* because an appeal in a related case involving Amazon was consolidated with Google’s appeal. The Ninth Circuit made a variety of other copyright rulings not relevant to this discussion.

transformative.”⁴ The court noted that in *Arriba Soft*, Arriba’s use “was transformative because Arriba’s use of the images served a different function than Kelly’s use – improving access to information on the Internet versus artistic expression. Although an image may have been created originally to serve an entertainment, aesthetic, or informative function, a search engine transforms the image into a pointer directing a user to a source of information.”⁵ Just as parody has transformative value because it provides social benefit by shedding light on an earlier work, “a search engine provides social benefit by incorporating an original work into a new work, namely an electronic reference tool.”⁶ In fact, the court went so far as to say that “a search engine may be more transformative than a parody,” the quintessential fair use, “because a search engine provides an entirely new use for the original work, while a parody typically has the same entertainment purpose as the original work.”⁷

The court observed that the fact that Google incorporated the entire Perfect 10 image into the search results does not diminish the transformative nature of Google’s use. “[E]ven making an exact copy of a work may be transformative so long as the copy serves a different function than the original work.”⁸ Further, when considering the third fair use factor, the amount and substantiality of the portion used, the Ninth Circuit reiterated its analysis of this issue in *Arriba Soft*: “it was necessary for Arriba to copy the entire image to allow users to recognize the image and decide whether to pursue more information about the image or the originating website. If Arriba only copied part of the

⁴ *Id.* at 1165.

⁵ *Id.*

⁶ *Id.*

⁷ *Id.*

⁸ *Id.*

image, it would be more difficult to identify it, thereby reducing the usefulness of the visual search engine.”⁹

Accordingly, the Ninth Circuit “conclude[d] that the significantly transformative nature of Google's search engine, particularly in light of its public benefit, outweighs Google's superseding and commercial uses of the thumbnails in this case.”¹⁰ The Ninth Circuit stated that in reaching this conclusion, it was mindful that the Supreme Court had stressed “the importance of analyzing fair use flexibly in light of new circumstances[,] ... especially during a period of rapid technological change.”¹¹ A few pages later, the Ninth Circuit added that

Google has put Perfect 10's thumbnail images (along with millions of other thumbnail images) to a use fundamentally different than the use intended by Perfect 10. In doing so, Google has provided a significant benefit to the public. Weighing this significant transformative use against the unproven use of Google's thumbnails for cell phone downloads, and considering the other fair use factors, all in light of the purpose of copyright, we conclude that Google's use of Perfect 10's thumbnails is a fair use.¹²

The Ninth Circuit made another important fair use holding. The district court's found that the cache copy made by a user's browser whenever he viewed a webpage is a fair use. The issue only arose because to prove secondary liability for Google, Perfect 10 needed to show that there was an underlying direct infringement by a third person. Among other possible direct infringements, Perfect 10 argued that users infringed its copyright in its images when they made temporary copies of these images in the random

⁹ *Id.* at 1167.

¹⁰ *Id.* at 1166.

¹¹ *Id.*, quoting *Sony Corp. v. Universal City Studios, Inc.*, 464 U.S. 417, 431-32 (1984), quoting H.R. Rep. No. 94-1476, p. 65-66 (1976), U.S. Code Cong. & Admin. News 1976, p. 5680.

¹² *Id.* at 1168.

access memory (RAM) of their computers while viewing these sites that posted these images without authorization. The Ninth Circuit agreed with the district court that the RAM copies were a fair use. It stated:

The copying function performed automatically by a user's computer to assist in accessing the Internet is a transformative use. Moreover, ... a cache copies no more than is necessary to assist the user in Internet use. It is designed to enhance an individual's computer use, not to supersede the copyright holders' exploitation of their works. Such automatic background copying has no more than a minimal effect on Perfect 10's rights, but a considerable public benefit.¹³

The *Perfect 10* court did not address Google's caching of the images gathered from websites, just its subsequent display of those images. This is because Perfect 10 focused on Google's display of the images, rather than its caching of them. Presumably, Perfect 10 concluded that under the *Arriba Soft* precedent, Google's caching of images from websites would be considered a fair use, and that it had a better shot at distinguishing *Arriba Soft* with respect to the display of the images because of the potential market for downloading images to mobile phones.

Perfect 10 represents a strong precedent for website archiving by a library. In *Perfect 10*, the court excused a commercial search engine's display of complete images. The court found that the use of a work for a different purpose – the location of information rather than entertainment – rendered the use transformative and therefore fair. Similarly, a library archives of websites has a different function from that for which the websites were created. The websites were “created originally to serve an entertainment, aesthetic, or informative function.” In contrast, the library website archives serves a preservation and scholarly function that provides a significant benefit to

¹³ *Id.* at 1169.

the public. The copying of entire websites is necessary for the accomplishment of this preservation and research function. And because the websites archived were distributed for free on the Internet, any market harm is completely speculative, as it was in *Perfect 10*.

II. *A.V. v. iParadigm* (2009)

iParadigm offered a plagiarism-detection service, Turnitin. School districts signed up with the service, and required students to turn their papers in to the service for comparison to papers (and journal articles) archived in the service's database. The new papers were then archived automatically, thereby increasing the size of iParadigm's database. When a student turned a paper in to iParadigm, she had to click on a license authorizing iParadigm to archive her paper. Several students required to submit their papers to iParadigm wrote on the papers that they were not authorizing the papers to be archived. When iParadigm nonetheless archived their papers, they sued iParadigm for infringing their copyrights. The district court ruled in favor of iParadigm, finding that 1) the students authorized iParadigm via the click-wrap license to archive their papers (notwithstanding the disclaimers on their papers); and 2) the archiving was a fair use.

The students appealed, and the Fourth Circuit affirmed the district court's fair use finding. The Fourth Circuit agreed with the district court that the commercial nature of iParadigm's use did not establish a per se bar to fair use. The district court correctly “determined that the commercial aspect was not significant in light of the transformative nature of iParadigm’s use.”¹⁴

The Fourth Circuit also rejected the plaintiffs’ argument that iParadigm’s use

¹⁴ *A.V. v. iParadigm*, 562 F.3d 630, 639 (4th Cir. 2009).

could not be transformative “because the archiving process does not add anything to the work – Turnitin merely stores the work unaltered in its entirety.”¹⁵ Relying on the Ninth Circuit’s decision in *Perfect 10 v. Amazon*, discussed above, the Fourth Circuit stated that

this argument is clearly misguided. The use of a copyrighted work need not alter or augment the work to be transformative. Rather, it can be transformative in function or purpose without altering or actually adding to the original work.¹⁶

The Fourth Circuit found that “the archiving was transformative because “iParadigm’s use of plaintiffs’ works had an entirely different function and purpose than the original works; the fact that there was no substantive alteration to the works does not preclude the use from being transformative in nature.”¹⁷

The plaintiffs next argued that even if a use has a transformative purpose, the use is not transformative if it fails to effect that purpose. Turnitin, they claimed, could not detect paraphrasing or verbatim copying in some cases. The court rejected this argument, too: “The question of whether a use is transformative does not rise or fall on whether the use perfectly achieves its intended purpose.”¹⁸ The Fourth Circuit concluded its discussion of the first fair use factor by observing that “iParadigm’s use of these works was completely unrelated to the expressive content and was instead aimed at detecting and discouraging plagiarism.”¹⁹

Thus, although iParadigm was a for-profit company, it copied entire unpublished works, and it did not alter them in any way, its archiving was fair because it repurposed the works it archived. The significance of the *iParadigm* decision lies in its application of

¹⁵ *Id.*

¹⁶ *Id.*

¹⁷ *Id.*

¹⁸ *Id.* at 640.

the reasoning of *Perfect 10* (and *Arriba Soft* before it) at the appellate level outside the Ninth Circuit. The conclusion that the fair use doctrine permits the archiving of works in a database for a new purpose cannot be dismissed as a Ninth Circuit aberration. Libraries outside of the Ninth Circuit can rely on the reasoning of *Perfect 10* and *iParadigm* to authorize their website archiving activities.

III. Patent and Trademark Office Opinion (2012)

In January 2012, the general counsel of the U.S. Patent and Trademark Office (USPTO) issued an opinion on the fair use status of the copying of non-patent literature (technical journal articles that demonstrate the state of knowledge – the “prior art” – in a field).²⁰ The opinion considered the copying and distribution of NPL in three situations:

- 1) USPTO copying NPL and providing those copies to applicants during the patent examination;
- 2) USPTO providing certified copies of entire patent application file histories, including copyrighted NPL, to members of public; and
- 3) Applicants making copies of NPL and submitting them to the USPTO during the patent examination process.

The USPTO obtains much of its NPL through licenses that permit it to make copies for use in examinations. However, the PTO occasionally makes copies of unlicensed NPL and provides these copies to applicants. The opinion applies to the unlicensed NPL.

The general counsel carefully applied the four factors in section 107 in the first

¹⁹ *Id.*

²⁰ See Bernard Knight, USPTO General Counsel, USPTO Position on Fair Use Copies of NPL Made in Patent Examination, January 19, 2012, http://www.uspto.gov/about/offices/ogc/USPTOPositiononFairUse_of_CopiesofNPLMadeinPatentExamination.pdf.

scenario, and concluded that the copying the NPL and providing the copies to the applicants constituted fair use. It then concluded that the analysis would be largely the same in the other two scenarios.

In its analysis of the first fair use factor in the first scenario (providing NPL to applicants), the general counsel found that “the USPTO is using the works for a non-commercial, governmental purpose, and is using the works not for their expressive content, but as evidence relating to the factual question of whether an invention is novel or non-obvious in view of the prior art as of a certain date.”²¹ Relying on *Perfect 10* and *iParadigm*, the general counsel found such a use to be transformative:

In providing applicants prior art NPL in the course of a patent examination, the USPTO is arguably using the copyrighted work for a new and different purpose than that for which it was created: the USPTO is using the NPL (or more accurately, just the relevant part of the NPL) to document, solely for the purposes of patent examination/prosecution, that certain features of the applicants’ claims are already in the prior art, or are obvious in view of the prior art. Under the principles discussed in the above-cited cases, the USPTO’s use of copyrighted NPL in the examination process could well be considered a “transformative” use for purposes of the first factor.²²

The general counsel also cited other “transformative use” cases where works were copied without alteration, including *Arriba Soft*, *Bill Graham Archives v. Dorling Kindersley*²³ and *Monge Maya Magazines, Inc.*²⁴

With respect to the third fair use factor, the general counsel stated that “[t]he

²¹ *Id.* at 3.

²² *Id.*

²³ 448 F.3d 605 (2d Cir. 2006)(finding use of copies of Grateful Dead posters in a biographical work about the Grateful Dead to be transformative).

²⁴ No. 09-5077, 2010 WL 3835053 (C.D. Cal. Sept. 30, 2010)(finding that magazine’s publication of secret wedding photos was transformative use where the photos were used to refute pop star’s denial of wedding).

USPTO makes an effort to limit copying to that which is relevant to the issue before the USPTO in examination. In many cases, however, an entire publication (i.e., an entire journal article) is considered relevant.”²⁵ The general counsel explains that “copying less than an entire work is often not an option for purposes of imparting the necessary information to patent applicants.”²⁶ On this basis, the general counsel found the third factor to be neutral.

In his discussion of the fourth factor, the general counsel stated that the USPTO’s provision of NPL to patent applicants would not impair the NPL’s marketability:

The NPL at issue has typically been published several years before the USPTO’s use and by the time of the USPTO’s use usually has only limited commercial value. In addition, by ... preventing any possibility of systematic infringement through access to USPTO’s databases ... the USPTO has taken steps to ensure that copies of NPL used in examination do not become freely available on the internet. Given that the only identifiable market for these works that the USPTO’s use might “impair” appears to be the market for the use of these works in patent examination itself, there does not seem to be any cognizable market impairment for purposes of the fourth factor.²⁷

The general counsel also viewed the second scenario, providing to the public copies of files with the included NPL, as

transformative in that work is not being used for its original purpose, but is instead being used in order to carry out the purposes of the patent system. Moreover, the file wrapper use is additionally transformative in the sense that the file wrapper as a whole becomes a legal document with a unique significance in patent litigation proceedings.²⁸

Finally, the general counsel considered photocopies submitted by patent applicants to the USPTO. Although the entity submitting the art is typically commercial, the submission of all

²⁵ USPTO Position at 4.

²⁶ *Id.*

²⁷ *Id.* at 5.

²⁸ *Id.*

known prior art is required by law. The applicant, therefore, is not exploiting the work, just using it pursuant to a legal requirement.

The USPTO is the executive branch agency with greatest expertise in intellectual property law. According to its website, “[t]he USPTO leads efforts to develop and strengthen both domestic and international property protection and advises the Secretary of Commerce, the President of the United States, and the Administration on patent, trademark, copyright, and copyright protection.”²⁹ The website further states that

the passage of the American Inventors Protection Act of 1999 (AIPA) (P.L. 106-113) set the stage for the USPTO to advise the President, through the Secretary of Commerce, and all Federal agencies, on national and international IP policy issues, including IP protection in other countries. USPTO is also authorized by the AIPA to provide guidance, conduct programs and studies, and otherwise interact with international IP offices and international intergovernmental organizations on matters involving the protection of intellectual property.³⁰

The general counsel of the agency with this IP-related mission relied heavily on *Perfect 10* and *iParadigm* to conclude that its photocopying and distribution of entire articles is a transformative use because the purpose of its use is different from the purpose for which the articles were written. Libraries should feel comfortable relying on these cases in a similar manner.

IV. ARL Code of Fair Use Best Practices (2012)

After reviewing judicial fair use decisions, Professor Michael Madison observed that the courts were

implicitly or explicitly, asking about habit, custom, and social context of the use, using what Madison termed a ‘pattern-oriented’ approach to fair use reasoning. If the use was normal in a community, and you could understand how it was different from the original market use, then judges

²⁹ <http://www.uspto.gov/ip/global/index.jsp>.

³⁰ *Id.*

typically decided for fair use.³¹

Stated differently, an important consideration influencing judges' decisions in fair use cases is "whether the user acted reasonably and in good faith in light of standards of accepted practice in his or her field."³²

Based on this insight, the Association of Research Libraries undertook an effort to "document[] the considered views of the library community about best practices in fair use, drawn from the actual practices and experience of the library community itself."³³ The resulting Code of Best Practices "identifies eight situations that represent the library community's current consensus about acceptable practices for the fair use of copyrighted materials and describes a carefully derived consensus within the library community about how those rights should apply in certain recurrent situations."³⁴ One of these eight situations is website archiving – collecting material posted on the World Wide Web and making it available.

The Code states that archives of websites "represent a unique contribution to knowledge and pose no significant risks for owners of either the sites in question or third-party material to which those sites refer. In the absence of such collections, important information is likely to be lost to scholarship."³⁵ The Code asserts that selecting and collecting material from the Internet is highly transformative:

The collecting library takes a historical snapshot of a dynamic and ephemeral object and places the collected impression of the site into a new context: a curated historical archive. Material posted to the Internet

³¹ Patricia Aufderheide and Peter Jaszi, *Reclaiming Fair Use* (2011) at 71.

³² Association of Research Libraries, *Code of Best Practices in Fair Use for Research Libraries* (January 2012) at 9.

³³ *Id.* at 3.

³⁴ *Id.*

³⁵ *Id.* at 26.

typically serves a time-limited purpose and targets a distinct network of users, while its library-held counterpart will document the site for a wide variety of patrons over time. A scholar perusing a collection of archived web pages on the Free Tibet movement, or examining the evolution of educational information on a communicable disease, seeks and encounters that material for a very different purpose than the creators originally intended.³⁶

The Code adds that

Preserving such work can also be considered strongly transformative in itself, separate from any way that future patrons may access it. Authors of online materials often have a specific objective and a particular audience in mind; libraries that collect this material serve a different and broader purpose and a different and broader network of users. Libraries collect not only for a wide range of purposes today, but also for unanticipated uses by future researchers.³⁷

The Code articulates the principle that “it is fair use to create topically based collections of websites and other material from the Internet and to make them available for scholarly use.”³⁸ The Code then offers limitations that should be observed to assure that the case for fair use is strong. These limitations are:

Captured material should be represented as it was captured, with appropriate information on mode of harvesting and date.

To the extent reasonably possible, the legal proprietors of the sites in question should be identified according to the prevailing conventions of attribution.

Libraries should provide copyright owners with a simple tool for registering objections to making items from such a collection available online, and respond to such objections promptly.³⁹

The Code next identifies enhancements that could further strengthen the case for fair use:

Claims of fair use relating to material posted with “bot exclusion” headers to ward off automatic harvesting may be stronger when the institution has

³⁶ *Id.*

³⁷ *Id.*

³⁸ *Id.* at 27.

³⁹ *Id.*

adopted and follows a consistent policy on this issue, taking into account the possible rationales for collecting Internet material and the nature of the material in question.

The more comprehensive a collection of web impressions in a given topic area is, the more persuasively the inclusion of any given item can be characterized as fair use.⁴⁰

The Code describes the library community's current consensus about acceptable practices for the fair use in the context of website archiving. A court likely will give weight to this description in assessing whether a library's archiving of a website constitutes a fair use.

Moreover, as stated at the conclusion of the 2006 memorandum, "library archiving of websites will take place in an environment where all the leading search engines, including Google, Yahoo, [Bing], and Ask, routinely cache millions of websites. A court will evaluate a library's ... fair use defense[] in the context of this industry practice."⁴¹ This industry practice is six years more entrenched now than in 2006, with the commercial search engines archiving billions of web pages.⁴² Furthermore, this industry practice has faced absolutely no legal challenge in the United States since the *Perfect 10* decision in 2007.

February 23, 2012

⁴⁰ *Id.*

⁴¹ Jonathan Band, "A New Day For Website Archiving: *Field v. Google* And *Parker v. Google*" (2006), arl.org, at 26. *See also* Jonathan Band, *Google and Fair Use*, 3 J. Bus. & Tech. L. 1 (2008).

⁴² <http://www.worldwidewebsite.com/>.

APPENDIX

A NEW DAY FOR WEBSITE ARCHIVING: *FIELD* v. *GOOGLE* AND *PARKER* v. *GOOGLE*

SUMMARY

As the World Wide Web has evolved into a major means of distributing information, librarians have recognized that archiving websites is an important element of cultural and historic preservation. However, the threat of copyright liability has prevented the systematic archiving of these sites by libraries. Fortunately, two recent district court decisions, *Field v. Google*, 412 F. Supp. 2d 1106 (D. Nev. 2006), and *Parker v. Google*, 422 F. Supp. 2d 492 (E.D. Pa. 2006), provide several viable legal theories for library website preservation.

The district court in *Field* identifies two strong defenses against copyright liability for libraries seeking to engage in the systematic archiving of websites: implied license and fair use. These defenses will be strongest with websites that do not employ “no-archive” meta-tags. The *Field* decision also suggests that the display of an archived site might not constitute a direct infringement by the library because it does not involve a volitional act by the library. The *Parker* decision could be understood as applying this volitional act theory to both the reproduction and the display of the website. In the absence of authority to the contrary, these two district court opinions from different circuits provide a solid basis for arguing that copyright law currently permits libraries to archive websites.

INTRODUCTION

The Supreme Court has observed that the Internet is “a unique and wholly new medium of worldwide communication.” *Reno v. ACLU*, 521 U.S. 844 (1997). It

“constitutes a vast platform from which to address and hear from a worldwide audience of millions of readers, viewers, researchers and buyers.” *Id.* at 853. The Court marveled at the “vast democratic fora of the Internet,” *id.* at 868, including thousands of newsgroups, “each serving to foster an exchange of information or opinion on a particular topic running the gamut from, say, the music of Wagner to Balkan politics to AIDS prevention to the Chicago Bulls.” *Id.* at 851. While the Internet has made more information more accessible to more users than any previous medium, it has a significant drawback: impermanence. Unlike a published book, which can remain accessible to users for centuries through libraries, the information on the World Wide Web is available only as long as it is stored on a computer attached to the Internet.

As the Internet has evolved into an increasingly important means of distributing government information, news, financial data, and artistic expression, librarians have become aware of the need to find a way to preserve this information of significant historical and cultural value. The vast size and dynamic nature of the Internet make website preservation an overwhelming technical task. But perhaps even more daunting than the technical challenge of gathering, storing, and indexing millions of constantly changing websites is the legal barrier of copyright. Much of the content of these websites falls within copyright’s protection, and libraries will not copy this content if doing so will subject them to legal liability. While many librarians believed that the fair use privilege, 17 U.S.C. 107, would permit website preservation, there was no case law sufficiently on point to provide librarians with enough certainty to justify proceeding with large-scale website archiving projects. However, the recent decisions in *Field v. Google*, 412 F. Supp. 2d 1106 (D. Nev. 2006), and *Parker v. Google*, 422 F. Supp. 2d 492 (E.D. Pa.

2006),⁴³ provide several viable legal theories for library website preservation.

FIELD v. GOOGLE

An attorney named Blake Field “decided to manufacture a claim for copyright infringement against Google in the hopes of making money from Google's standard practice” of caching the content it spiders and presenting users with links to the cached copies. 412 F. Supp. 2d at 1113. He wrote 51 short stories and posted them on his website. Field was aware that he could use a “bot exclusion” header in a robot.txt file, which would instruct Google’s software that gathers material from the Web – the “Googlebot” -- not to copy the site into Google’s search index. He also was aware that he could include a "no archive" meta-tag in the robot.txt file, in which case Google would copy the site into its search index, but it would not keep a cached copy of the site, nor provide users with links to the cache. Instead, he set the permissions within the robot.txt file to allow all search engine “bots,” including Google’s, to visit and index the site, and he intentionally did not employ a “no-archive” meta-tag. Detecting neither a bot exclusion header nor a “no archive” meta-tag, Google crawled Field’s site, included his stories in its search index, and provided links to cached copies of the stories.

Field sued Google for copyright infringement. He argued that the Google Cache feature, by allowing Google users to link to archived copies of his stories indexed by Google, infringed his copyrights. Field did not argue that Google infringed by virtue of copying his stories into its index in the first place. He presumably was familiar with *Kelly v. Arriba Soft*, 336 F.3d 811 (9th Cir. 2003), where the Ninth Circuit found that a search engine’s storage and display of thumbnail-sized images copied from websites, with links

⁴³ *Parker* currently is on appeal to the Third Circuit.

to the full-sized images on the original websites, constituted a fair use. Since *Kelly* is binding precedent on the federal court in Nevada, Field likely assumed that a court would treat Google's copying and storing of his stories as a fair use. By centering his complaint on Google's serving the cache copy of his stories to users, rather than on Google's act of making and storing the cache copy, Field probably hoped to distinguish his case from *Kelly*. The *Kelly* court's fair use ruling was based in large measure on its conclusion that a display of a thumbnail sized image did not supersede the display of a full-sized image, and that a user still had to go to Kelly's site to see the full-sized image. Here, by contrast, the stories served from the Google cache were identical to the stories displayed on Field's website.

Nonetheless, the district court granted summary judgment in favor of Google on five independent bases, several of which have direct applicability to website archiving by libraries.

A. No Volitional Act. The *Field* court found that Google did not directly infringe copyright by serving a webpage from the Google Cache, because Google did not engage in any volitional activity with respect to the serving of the cached webpage. Rather, a user initiated the serving of the cached copy by clicking on the cache link. Then, Google's servers displayed the copy to the user by an automatic process. In reaching this conclusion, the district court relied on *Religious Tech. Ctr. v. Netcom On-Line Commc'n Servs., Inc.*, 907 F. Supp. 1361, 1369-70 (N.D. Cal. 1995), where the court held that direct infringement requires a volitional act by the defendant, and that the automated copying by machines owned by the defendant was insufficient to trigger liability. The *Field* court also relied on *CoStar Group, Inc. v. LoopNet, Inc.*, 373 F.3d 544, 555 (4th Cir.

2004), where the Fourth Circuit stated that “[a]greeing with the analysis in *Netcom*, we hold that the automatic copying, storage, and transmission of copyrighted materials, when instigated by others, does not render an ISP strictly liable for copyright infringement....”

The *Field* court only found that Google's display of the cache copy did not involve a volitional act by Google. It did not consider whether Google's caching of the copy by sending out the Googlebot to crawl and copy websites involved volition because those acts were not the subject of this lawsuit. Thus, the *Field* court's holding concerning volitional acts protects libraries only with respect to their providing access to archived websites in response to user requests. Since the decision did not reach Google's caching of websites, there is no way to know whether the *Field* court would find volition in a library's automated copying and storing of websites. Still, at the very least, the volitional act theory could protect libraries in connection to an important part of the website archiving function.

B. Implied License. More helpful to libraries is the court's holding that Field's posting an "allow all" robot.txt header and then failing to set a "no archive" meta-tag indicated that he impliedly licensed search engines to permit users to access the cached copies. The court found that “the ‘no-archive’ meta-tag is a highly publicized and well-known industry standard” of which Field was aware. 412 F. Supp. 2d at 1116. The court further found that “Google reasonably interpreted the absence of meta-tags as permission to present ‘Cached’ links to the pages of Field's site.” *Id.* On this basis, the court concluded that Field granted Google an implied license for this use.

This ruling suggests that libraries can argue that website operators have authorized libraries both to archive and display their sites if the operators fail to employ a

“no archive” meta-tag. This implied license defense would work best with respect to actions a library takes prior to receiving a cease-and-desist letter from the website operator. However, once the library received such a letter, it could not argue that it had an implied license to continue displaying the archived website. Nonetheless, the library might succeed in convincing a court that since it reasonably relied on the absence of the “no-archive” meta-tag when it initially cached the website, it should be permitted to continue to display the cache. While a court might permit the library to continue displaying the archived website, it probably would not permit the library to archive updates to the website.

It should be noted that this implied license defense has a significant practical limitation: many websites with valuable content, such as government websites, often employ “no-archive” meta-tags or bot exclusion headers. A library obviously could not claim an implied license to archive such websites.

C. Fair Use. The *Field* court extended *Kelly* from the display of thumbnail images to the display of the complete text of web sites. Considering the first fair use factor, the purpose and character of the use, the *Field* court found that the display of cached copies served at least three different purposes from the original work and therefore was transformative and did not supersede the original.

First, and most relevant to library archiving,

Google’s ‘Cached’ links allow users to view pages that the user cannot, for whatever reason, access directly. A Web page can become inaccessible to Internet users because of transmission problems, because nations or service providers seek to censor certain information, because too many users are trying to access the same page at the same time, or because the page has been removed from its original location. In each case, users who request access to the material from the inaccessible site are still able to access an archival copy of the page via the ‘Cached’ link in Google’s

search results. Google's users, including those in academia, describe this functionality as highly valuable.

Id. at 1111 (citations omitted). The court further explained that this archiving function also benefited the original website publishers because it allowed them "to recover copies of their own sites that might otherwise have been lost due to computer problems." *Id.* at 1112 (citation omitted). The court observed that the "Internet is replete with references from academics, researchers, journalists, and site owners praising Google's cache" for providing this archiving functionality. *Id.* at 1118. When Google provides access to an otherwise inaccessible website, "Google's archival copy of a work obviously does not substitute for the original." *Id.*

Second, the cache copy allows users to detect changes to a website, which can be significant for political, educational, and legal reasons. A user can identify changes only by comparing the current website with the original archived by Google. Third, the cache copy allows a user to understand why a page is responsive to a query, because the queried term is highlighted in the cache copy.

In sum, even though the cache copy is identical to the original, the court found the display of the cache copy to be transformative because it "serves different and socially important purposes" from the original. *Id.* at 1119. If Google's archiving is transformative, then a library's preservation of websites is *a fortiori* transformative.

Field argued that Google's commercial objectives cut against fair use. The court dismissed this argument by stressing that there was no evidence that Google profited from its use of Field's stories: "Field's works were among *billions* of works in Google's database." *Id.* at 1120. A library could make an even stronger first factor argument than Google because its archiving obviously serves a non-commercial preservation purpose.

Turning to the second factor, the nature of the copyright works, the court noted that Field had made his works “available to the widest possible audience for free” through his website. *Id.* The court evidently concluded that this sort of use by its owner implied that the work was not that creative and not that deserving of protection.

When discussing the third factor, the amount of the work used, the district court observed that the “multiple and socially valuable purposes” of the cache could not be achieved by the display of only portions of web pages:

Without allowing access to the whole of a web page, the Google Cached link cannot assist Web users (and content owners) by offering access to pages that are otherwise not available. Nor could use of less than the whole page assist in the archival or comparative purposes of Google’s ‘Cached’ links. Finally, Google’s offering of highlighted search terms in cached copies of Web pages would not allow users to understand why a Web page was deemed germane if less than the whole Web page were provided.

Id. at 1121. In sum, Google “displayed no more of the works than is necessary” to accomplish its purpose. *Id.*

With respect to the fourth factor, the market impact of the cache copy, the court noted that Field made his works available for free on the Internet, and never received any revenue for them from any source. Since there was no evidence of any market for Field’s works, there could be no harm to that market. Further, the court found that “there is no evidence ... of any market for licensing search engines the right to allow access to Web pages though ‘Cached’ links, or evidence that one is likely to develop.” *Id.* at 1122. To the contrary, the court observed that

[t]here is compelling evidence that site owners would not demand payment for this use of their works. Notwithstanding Google’s long-standing display of ‘Cached’ links and the well-known industry standard protocols for instructing search engines not to display them, the owners of literally billions of Web pages choose to permit such links to be displayed.

Id. In the absence of a market for the licensing of “caching rights,” Google’s caching of his stories did not deprive him of any revenue. So, too, a library’s preservation of a website typically would not deprive the website operator of any revenue. Website archiving could conceivably affect the market for a commercial website by diverting traffic from that site, but presumably such a site would use either a bot-exclusion header or a “no-archive” meta-tag. If such an operator failed to use a “no-archive” meta-tag, he would have difficulty convincing a court that the site’s content had any commercial value that archiving could harm.

The *Field* court then considered a fifth, non-statutory, fair use factor: “whether an alleged infringer has acted in good faith.” *Id.* The court found that Google honored industry standard protocols such as bot exclusion headers and “no-archive” meta-tags. Google’s website provided website operators with explanations on how to deploy these protocols. Google also provided an automated mechanism for a website operator to prevent the further display of a cached website after Google cached it. Finally, in this instance, Google disabled the links to the cache as soon as Field filed his complaint. Hence, the court concluded that Google acted in good faith.

While libraries could easily emulate all of Google’s caching policies, they might be reluctant to provide an automated mechanism for website operators to prevent access to archived material. Because a library understandably would oppose outside interference with its preservation decisions, a court would likely still conclude that a library acted in good faith even if it did not adopt this particular feature.

However, a library might have a more difficulty using *Field*’s fifth factor as a precedent if the library ignored “no-archive” meta-tags. And, as noted above, a library’s

position on market harm is stronger if it respects “no-archive” meta-tags. Although a library may still prevail in its fair use defense if it archived a website in contravention of a “no-archive” meta-tag, the library is more likely to succeed in cases involving websites that did not employ “no-archive” meta-tags.

4. Digital Millennium Copyright Act. The *Field* court held that the Section 512(b) caching "safe harbor" for online service providers in the Digital Millennium Copyright Act applied to Google’s caching of website content. Section 512(b) limits a service provider's liability for the intermediate and temporary storage of material on its system if (a) the material is made available online by a person other than the service provider; (b) the material is transmitted from the person who made the material available online to another person at the direction of that “other person;” and (c) the storage is carried out through an automatic technical process for the purpose of making the material available to users. The *Field* court concluded that the "other person" in (b) could be Google itself: "Field transmitted the material in question, the pages of his Web site, to Google's Googlebot at Google's request. Google is a person other than Field. Thus, Google's cache meets the requirements of Section 512(b)(1)(B)." *Id.* at 1124.

The system caching safe harbor has typically been understood to protect a service provider that cached material in the course of the material’s transmission from a website to an end-user. It was assumed that for purposes of Section 512(b), the “other person” was the end-user, and not the service provider. At the same time, the *Field* court's interpretation is not inconsistent with the plain language of the statute; the service provider is a person other than the person who made the material available online. Thus, on its face, Section 512(b) could apply to website archiving.

Nonetheless, Section 512(b) might be of limited utility to libraries. Field argued that Section 512(b) did not apply because Google’s storage was not “intermediate and temporary.” The court noted that Google’s cache was “intermediate” because it was “a repository of material that operates between the individual posting the information, and the end-user requesting it.” *Id.* at 1124. With respect to “temporary,” the court noted that the Ninth Circuit in *Ellison v. Robertson*, 357 F.3d 1072 (9th Cir. 2004), considered AOL’s storage of a Usenet posting for 14 days to be “transient” for purposes of 17 U.S.C. § 512(a). Accordingly, the *Field* court found that Google’s cache of material for 20 days to be “temporary” under Section 512(b). While a library might succeed in arguing that its archival copy is intermediate because its objective is providing end-users with access to the material, a library presumably will want to archive websites for far longer than 20 days. Courts probably will not treat a copy that resides in a computer server for years as “temporary.”

While Section 512(b) does not appear to provide libraries with adequate protection for website archiving, the *Field* court’s holdings with respect to implied license and fair use provide libraries with two viable defenses. Additionally, although the *Field* court’s ruling concerning volitional acts might not help libraries with respect to the archiving of websites, it could assist libraries with the display of archived sites.⁴⁴

⁴⁴ The *Field* court also ruled that Field was estopped from asserting a copyright claim because he induced Google to infringe by using software code that invited Google to scan and cache his website, but then intentionally failed to instruct Google not to serve the cached copies. This defense turns on the peculiar facts of this case, and would not assist libraries when they sought to archive most websites.

PARKER v. GOOGLE

Just two months after the *Field* decision, a federal district court in Pennsylvania issued a decision that built on some of *Field's* holdings. Parker represented himself in the case, so the court had difficulty understanding exactly what he was claiming. As a result, the opinion is unclear on some points of fact and law. Parker, an author, posted his writings on his website. He also posted them to Usenet groups. Google automatically archived his Usenet postings. Additionally, Google cached and displayed the content of his website. Finally, when Google presented a list of hyperlinks in response to search queries, it excerpted portions of his website.

Parker claimed that these Google actions constituted direct copyright infringement. The court granted Google's motion to dismiss. Relying on *CoStar* and *Netcom*, the court found that Google had not engaged in a volitional act:

When an ISP automatically and temporarily stores data without human intervention so that the system can operate and transmit data to its users, the necessary element of volition is missing. The automatic activity of Google's search engine is analogous. It is clear that Google's automatic archiving of USENET postings and excerpting of websites in its results to users' queries do not include the necessary volitional element to constitute direct copyright infringement.

422 F. Supp. 2d at 497. Here, the court appears to suggest that the archiving of a website, if performed automatically and without human intervention, does not involve a volitional act. Thus, the *Parker* court seems to go beyond the *Field* court. While the *Field* court ruled that the display of an archived work did not involve a volitional act by Google, the *Parker* court indicated that the initial archiving did not involve a volitional act either.⁴⁵

⁴⁵ Citing *Field* as authority, the *Parker* court also found that caching safe harbor at Section 512(b) of the DMCA sheltered Google's caching of web pages as a means of

Even if a library does not engage in a volitional act when archiving a website or displaying it, could it nonetheless be secondarily liable for the infringement made by the user when he accesses the site? After all, the user makes a copy in the random access memory of his computer, and he might also print out a hard copy. *Parker* suggests that a library would not meet the knowledge requirement for contributory infringement. It found that Google never had the requisite knowledge of the copying facilitated by its automatic systems. So, too, a library would not have knowledge that a specific use made by a specific user was infringing, until after the website operator informed the library that it did not want the library to archive and display its website. If the website did not employ a bot exclusion header or a “no-archive” meta-tag, the library could reasonably assume that the website operator was permitting not only the library’s reproduction and display of the site, but also the user’s RAM and hard copies.

Parker also suggests that a library would not satisfy the criteria for vicarious liability. The *Parker* court did not find any evidence of financial benefit resulting from any specific act of infringement. The court quoted with approval Nimmer’s language that “[l]arge commercial ISPs derive insufficient revenue from isolated infringing bits, in the context of the billions of bits that cross their servers, to characterize them as financially benefiting from the conduct of which complaint is made.” *Id.* at 500. The court also noted that in *Ellison*, the “record lack[ed] evidence that AOL attracted or retained subscriptions because of the infringement or lost subscriptions because of AOL’s eventual obstruction of the infringement.” *Id.* If the *Parker* court found that Google did not benefit financially

indexing websites and producing results to search queries. The *Parker* court did not address the meaning of the term “temporary storage.”

from the infringing conduct of its users, it is hard to imagine that any court would find that a library benefits financially from its users' infringing use of its archive.

Finally, and perhaps most importantly, a library can be secondarily liable for its users' activities only if those activities are themselves infringing. In the overwhelming majority of situations, the user of a library's website archive would have a strong fair use argument excusing her from copyright liability.

CONCLUSION

The *Field* court provides two powerful defenses against copyright liability for libraries considering the systematic preservation of websites: implied license and fair use. These defenses will be strongest with websites that do not employ "no-archive" meta-tags. *Field* also suggests that the display of an archived site might not constitute a direct infringement by the library because the display does not involve a volitional act by the library. The *Parker* decision could be viewed as applying this theory to both the reproduction and the display of the website. In the absence of authority to the contrary, these two district court opinions from different circuits provide a solid basis for arguing that copyright law currently permits libraries to archive websites.

Moreover, library archiving of websites will take place in an environment where all the leading search engines, including Google, Yahoo, MSN, and Ask, routinely cache millions of websites. A court will evaluate a library's implied license and fair use defenses the context of this industry practice.

Originally published in 2006